AD NUMBER
AD886376
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; 30 JUN 1971. Other requests shall be referred to Air Force Technical Applications Center, Patrick AFB, FL.
AUTHORITY
AFTAC ltr, 1 May 1972

AD886376

Computation of Multiple-Event Probabilities,

TELEDYNE GEOTECH ALEXANDRIA VA

30 JUN 1971

COMPUTATION OF MULTIPLE-EVENT PROBABILITIES

SEISMIC DATA LABORATORY REPORT NO. 277

AFTAC Project No.:	VELA T/2706
Project Title:	Seismic Data Laboratory
ARPA Order No.:	1714
ARPA Program Code No.:	2F-10
Name of Contractor:	TELEDYNE GEOTECH
Contract No.:	F33657-72-C-0009
Date of Contract:	01 July 1971
Amount of Contract:	$ 1,290,000
Contract Expiration Date:	30 June 1972
Project Manager:	Royal A. Hartenberger (703) 836-7647

P. O. Box 334, Alexandria, Virginia

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) TELEDYNE GEOTECH ALEXANDRIA, VIRGINIA	2a REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP

3 REPORT TITLE

COMPUTATION OF MULTIPLE-EVENT PROBABILITIES

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific

5 AUTHOR(S) (Last name, first name, initial)

Wirth, Mark H.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO OF REFS
6/30/71	14	3

8a CONTRACT OR GRANT NO. [redacted] F33657-72-C-0009 b. PROJECT NO. VELA T/0706 c. ARPA Order No. 624 d. ARPA Program Code No. 9F10	9a ORIGINATOR'S REPORT NUMBER(S) 277 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency Nuclear Monitoring Research Office Washington, D. C.

13 ABSTRACT

A new, simple and efficient algorithm is described which computes the probability of occurrence of at least (or exactly) k out of N events, given the individual event probabilities. A comparison is made with two previous methods. Computer program listings are given.

14 KEY WORDS

Probabilities
Computation
Algorithms

Unclassified

Security Classification

ABSTRACT

A new, simple and efficient algorithm is described which computes the probability of occurrence of at least (or exactly) k out of N events, given the individual event probabilities. A comparison is made with two previous methods. Computer program listings are given.

TABLE OF CONTENTS

THE ALGORITHM

It is desired to compute the probability of at least k out of N events occurring, given the individual event probabilities p_i. This can be obtained by summation from the probability of exactly k events, P(k), i.e.

$$P(\geq k) = \sum_{i=k}^{N} P(i) \qquad (1)$$

Thus we need only compute the probability of exactly k events.

The algorithm is based on the observation that given two groups of events, with their associated probabilities of occurrence, the probability of exactly k total out of the two groups can be found by

$$P\left(\binom{N_1+N_2}{k} \middle| S_1 \cup S_2\right) = \sum_i P\left(\binom{N_1}{i} \middle| S_1\right) P\left(\binom{N_2}{k-i} \middle| S_2\right) \qquad (2)$$

where by $P(\binom{N}{k}|S)$ we denote the probability of exactly k out of N events from the set S. The index i is subject to the limitations $0 \leq i \leq N_1$ and $0 \leq k-i \leq N_2$, i.e. sum over $i = \max(0,k-N_2),\ldots,\min(k,N_1)$. Note that $P(\binom{1}{1}|S_j) = p_j$, and $P(\binom{1}{0}|S_j) = 1-p_j$, by definition. Thus we are initially given the probabilities for groups of size 1. Using the combining rule (2), we can recursively compute the probabilities for groups of size 2,4,8,...,N. Since (2) works for disparate group sizes, no problem is caused by the odd groups that result when N is not a

power of two. An odd group without a mate to pair with at any stage is simply passed on unchanged to the next stage: at a later stage it will be paired with a group having a different number of elements. A schematic illustration of the operation of the algorithm for N = 5:

```
input:          11111
               0101010101
               012-012-01
               01234---01
               012345----
output:         12345
```

where the numerals indicate the index of P(k) and the positions indicate the relative positions in an array of storage, i.e. each "2" in the diagram represents the location of a probability of exactly two events, etc. Dashes indicate blanks (unused locations). The final summation is not shown.

A listing of a computer subroutine (NETPROB III) employing the algorithm is given in the Appendix. The restriction $N \leq 50$ is completely arbitrary for this routine and derives from the requirement of at least 4N locations in the working array. The routine can be modified to output the probability of exactly k events by eliminating the final accumulation (statement 200). Note that the $P(\geq k)$ are computed for all k at once by this algorithm (k = 1,...,N).

TIMING

It is easily shown that the combining operation (2) requires $\sim M^2/4$ operations for $M = N_1 + N_2 = 2,4,8,\ldots,N$. Thus the total time should be proportional to

$$\frac{4}{4}\cdot\frac{N}{2} + \frac{16}{4}\cdot\frac{N}{4} + \frac{64}{4}\cdot\frac{N}{8} + \cdots + \frac{N^2}{4}\cdot 1$$

$$= \frac{N}{2}\left(1+2+4+\cdots+\frac{N}{2}\right) = \frac{N(N-1)}{2} \sim N^2$$

Timing tests on a CDC 1604B computer gave a timing curve

$$T \approx .08\,(N+25)N \quad \text{msec.}$$

REMARKS

Any N^2 process for which a combining rule analogous to (2) can be given requiring only M operations, where M is the total number of elements in the two groups, can be converted to an $N \log_2 N$ process by this binary subdivision algorithm. In fact, the author has used a virtually identical subdivision routine as the outer routine for a fast sort package. (Merge-ordering of two groups which are in themselves ordered can be easily done in no more than M operations.) Thus this algorithm should be useful as a general procedure for speeding many N^2 calculations.

The algorithm was applied in this case in order to avoid certain difficulties mentioned in the next section, not out of considerations of speed. Nevertheless, it is interesting to note that for large N a more efficient combining rule than (2) can be used. With the natural identification of $P(\binom{N}{k}|S)$ as the elements of an array, (2) is formally equivalent to a convolution. Well-known convolution techniques using the Fast Fourier Transform could be used to reduce the total number of computations to the order of $N(\log_2 N)^2$. Due to the fact that the FFT is more efficient only for rather large N, coupled with the necessity of padding to avoid wraparound, this approach was not thought desirable here. For extremely large N, it might hold some merit, but in that case limiting approximations might be more appropriate anyway.

COMPARISON WITH PREVIOUS METHODS

Feller (1950) gives a formal solution for the probability of exactly k events

$$P(k) = \sum_{j=k}^{N} (-1)^{j-k} \binom{j}{k} E_j \tag{3}$$

where

$$E_1 \equiv \sum_{i=1}^{N} p_i$$

$$E_2 \equiv \sum_{i<j} p_i p_j \tag{4}$$

$$E_3 \equiv \sum_{i<j<k} p_i p_j p_k$$

etc.

Booker (1965), in order to avoid the computation in (4), transformed this result into polynomials in S_j, where

$$S_j \equiv \sum_{i=1}^{N} \left[p_i/(1-p_i)\right]^j \tag{5}$$

with coefficients related to the Stirling numbers. His method may be useful for calculating only a few of the P(k), for small k, but if all the P(k) are to be calculated, then it will be quite slow, even assuming that all the coefficients are either available or may be recursively calculated (not an easy task in itself). In addition, the method will not work very well for p_i close to 1, due to the division in (5), nor is it obvious that his alternating series are numerically well-conditioned.

This author (Wirth, 1970) showed that all of the E_j could be computed at once recursively in approximately $N^2/2$ operations, using the scheme

$$\begin{array}{l}
E_1=p_1,\ E_2=p_2E_1,\ E_3=p_3E_2,\ \ldots,\ E_N=p_NE_{N-1} \\
E_1=E_1+p_2,\ E_2=E_2+p_3E_1,\ E_3=E_3+p_4E_2,\ \ldots,\ E_{N-1}=E_{N-1}+p_NE_{N-2} \\
E_1=E_1+p_3,\ E_2=E_2+p_4E_1,\ E_3=E_3+p_5E_2,\ \ldots,\ E_{N-2}=E_{N-2}+p_NE_{N-3} \\
\vdots \\
E_1=E_1+p_{N-1},\ E_2=E_2+p_NE_1 \\
E_1=E_1+p_N
\end{array}$$

(read like print, left to right, top to bottom). With this method, all of the P(k) may be calculated at once in $\sim N^2$ operations. The resulting program (NETPROB I, Appendix) is remarkably simple, but suffers from a

major limitation: for all p_i close to 1, the terms in (3) can become very large and truncation effects can become serious. The author showed that the largest term will be

$$\max \binom{N}{j}\binom{j}{k} \approx \frac{N!}{\left[(N/3)!\right]^3} \approx 3^N/N$$

For N = 20 this is ~2 x 10^8 and the worst case accuracy obtainable with CDC 1604 36-bit floating-point mantissas is only ~1/2%. Total loss of significance occurs for N = 27. Use of double-precision would allow the range of N to be extended, but is extremely time-consuming. Nevertheless, for small N the program is eminently satisfactory. (The recursion for the E_j should also be useful for reconstructing a polynomial from its roots.)

The method described in the first part of the paper, however, suffers from no such limitation and is, in addition, slightly faster. Thus it appears to be distinctly preferable. Tests performed by comparing NETPROB III against a double-precision version of NETPROB I for all $N \leq 40$ showed excellent agreement and no loss of significance.

REFERENCES

Booker, A., 1965, Numerical evaluation of symmetric polynomials: J. Assoc. Comp. Mach., v. 12, p. 90-94.

Feller, W., 1950, An introduction to probability theory and its applications: Wiley and Sons, v. 1.

Wirth, M.H., 1970, Estimation of network detection and location capability: Seismic Data Laboratory Memorandum, Teledyne Geotech, Alexandria, Virginia.

APPENDIX

Computer subroutines are written in FORTRAN-63, a programming language of the CDC 1604B computer.

```
      SUBROUTINE NEIPROB( N,P, PN )                                       III
      DIMENSION  P(N), PN(N), W(200)
C     PN(K) = PROB. OF AT LEAST K OUT OF N EVENTS WHEN INDIVIDUAL EVENTS
C     HAVE PROB. P(I).    N <= 50                                   M. WIRTH
C
      II = 2
      DO 5 I = 1,N
      W(II) = P(I)              $     W(II-1) = 1.- P(I)
    5 II = II + 2
      NING = 2                  $     LOC = NT2 = 2 * N
   10 NING2 = NING              $     N2 = NING2 / 2
      NING = NING * 2           $     NNG = NT2 - NING
      LOC2 = LOC                $     LOC = NT2 - LOC
      IF( NNG ) 100,100,20
   20 DO 30 IB = 1,NNG,NING
      L = LOC + IB
      CALL PROBK( N2,W(L),W(L+NING2),   NING2,W(LOC2+IB) )
   30 II = IB
      IB = II + NING            $     NR = NT2+1-IB
      IF( NR.GT.NING2 ) 40,50
   40 L = LOC + IB
      CALL PROBK( N2,W(L),W(L+NING2),   NR/2,W(LOC2+IB) )
      GO TO  10
   50 DO 60 I = IB,NT2
   60 W(LOC2+I) = W(LOC+I)
      GO TO  10
  100 CALL PROBK( N2,W(LOC+1),W(LOC+NING2+1), N,W(LOC2+1) )
      K = N                     $     PN(K) = W(LOC2+K+1)
      DO 200 I = 2,N
      K = K - 1
  200 PN(K) = PN(K+1) + W(LOC2+K+1)
      RETURN
      END
```

```
      SUBROUTINE PROBK( N2,P2,P1, NT,PT )
      DIMENSION  P1(N2), P2(N2), PT(NT)
C     PT(K+1) = PROB. OF EXACTLY K OUT OF NT EVENTS TOTAL FROM TWO
C     GROUPS, WITH P1,P2(I+1) = PROB. OF EXACTLY I EVENTS OUT OF N1,N2.
C
      N1 = NT - N2        $     IS = 0
      PT = P1 * P2
      DO 100 K = 1,NT
      PT(K+1) = 0.
      IF( K.GT.N1 ) 20,10
   10 IL = K              $     GO TO  40
   20 IL = N1
      IF( K.GT.N2 ) 30,40
   30 IS = K - N2
   40 DO 100 I = IS,IL
  100 PT(K+1) = PT(K+1) + P1(I+1)*P2(K-I+1)
      RETURN
      END
```

```
      SUBROUTINE NETPROB( N,P, PNET )
      DIMENSION P(N), PNET(N), E(20)
C     PNET(K) = PROB. OF AT LEAST K OUT OF N EVENTS WHEN INDIVIDUAL
C     EVENTS HAVE PROB. P(I).    N S 20
C                                                            M. WIRTH
      CALL ERASE( N,E )
      DO 10 K = 1,N
      E = E + P(K)          $     JM = N - K
      DO 10 J = 1,JM
   10 E(J+1) = E(J+1) + P(K+J)*E(J)
      PNET(N) = E(N)        $     NM = N - 1
      DO 30 K = 1,NM
      KR = N - K            $     PNET(KR) = E(KR)
      B = KR                $     C = 0.        $     A = 1.     $     KP = KR+1
      DO 20 J = KP,N
      B = B + 1.            $     C = C + 1.    $     A = -A * B/C
   20 PNET(KR) = PNET(KR) - A*E(J)
   30 PNET(KR) = PNET(KR) + PNET(KR+1)
      RETURN
      END
```